Statistical Analysis and Visualization of Medical Instruction Patterns in Multimedia-Rich Electronic Medical Records

Miwa Sugitani[1)], Ryosuke Matsuo[1)], Tomoyoshi Yamazaki[1)], Kenji Araki[1)],

Masato Oguchi[1)], Haruo Yokota[2)], **Hieu Hanh Le**[1)]

[1)] *Ochanomizu University, Japan*

[2)] *Josai University, Japan*

Research Background

Electronic Health Record (EHR) Adoption

- EHRs are spreading, primarily in large hospitals and expanding to medium and small hospitals
- The aggregation and secondary use of large volumes of **multimedia-rich data** are increasing
  - Medical order, images, natural language, specimen test results, etc.

Importance of EHR Data Analysis

- Analyzing frequent medical orders and clinical pathways contributes to the standardization and improvement of clinical processes, as well as enhancing diagnostic accuracy
  - Clinical Pathway: A standardized, chronological plan for the clinical process for a specific disease or treatment

ex. admission → surgery → test → drug → discharge

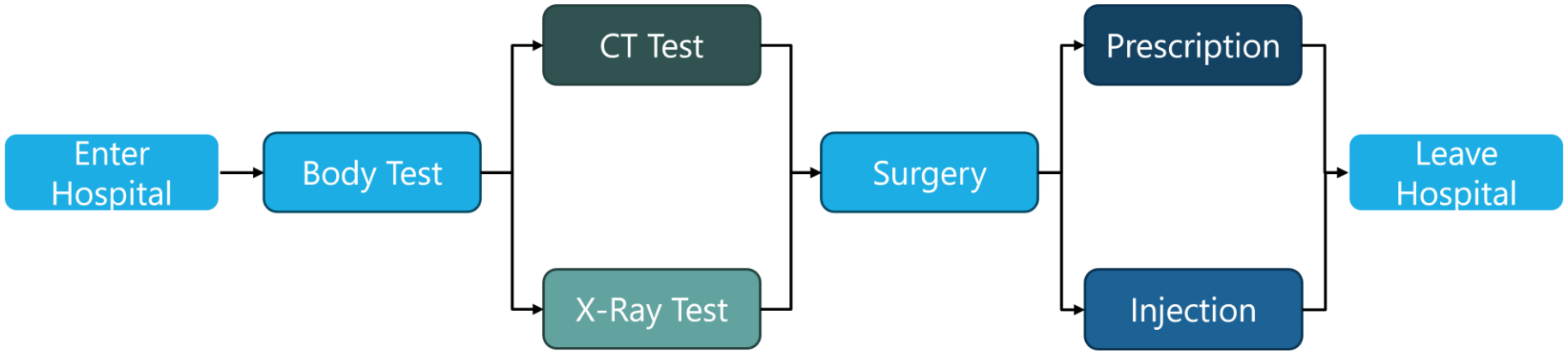
Personalized Treatment

- Appropriate medical orders can differ for the same disease due to patient age and health status.
- The medical field anticipates leveraging past treatment patterns and statistical information.

Related Works (2/2)

Frequent Medical Order Pattern Analysis from Single-Institution Data

- Visualize the generated frequent pattern as a Sequence Variant (SV) [Honda+. 2018].
  - SV is the extension of a sequence with branches
- Evaluate the reasons why variants might appear using multivariate analysis [Le+, 2019].
  - Test results, medical background, age, etc.

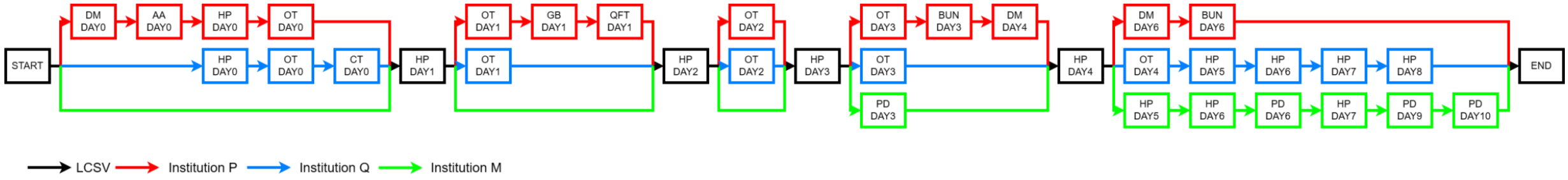


An SV example

Related Works (2/2)

Frequent Medical Order Pattern Analysis across Multiple-Institution Data

- A method to analyze transitions in differences between medical orders for COVID-19 treatment [Zhao+, 2023].
- Work has also analyzed EHR data related to COVID-19 across multiple institutions, extracting and clustering frequent instruction patterns based on sequence similarity [Le+, 2024].



These studies were limited to a single disease without statistical information in the extracted patterns

Contributions

1. **Dealing with multiple diseases across multiple medical institutions**
   - Extract frequent medical order patterns for each disease using anonymized EHR data from multiple medical institutions.

2. **Extracting useful statistical information for understanding differences in generated patterns**
   - Calculate statistical information, like test results, related to these frequent medical order patterns

3. **Developing a functional visualization tool for better understanding the outputs**
   - Allow for exploring branching in frequent medical order patterns while reviewing statistical information.

Proposed Method

1

Step 1: Medical Order Sequence Construction

Medical order sequences are created by arranging each patient's medical orders chronologically.

2

Step 2: Frequent Medical Order Pattern Extraction with Statistical Information

From these sequences, frequent medical order patterns are extracted for each disease. Statistical information, such as mean and median, is extracted for the medication and tests included in these frequent medical order patterns.

3

Step 3: Calculation of Abnormal Test Result Occurrence Rate

The occurrence rate of abnormal test results is calculated based on whether they are included in the frequent medical order patterns.

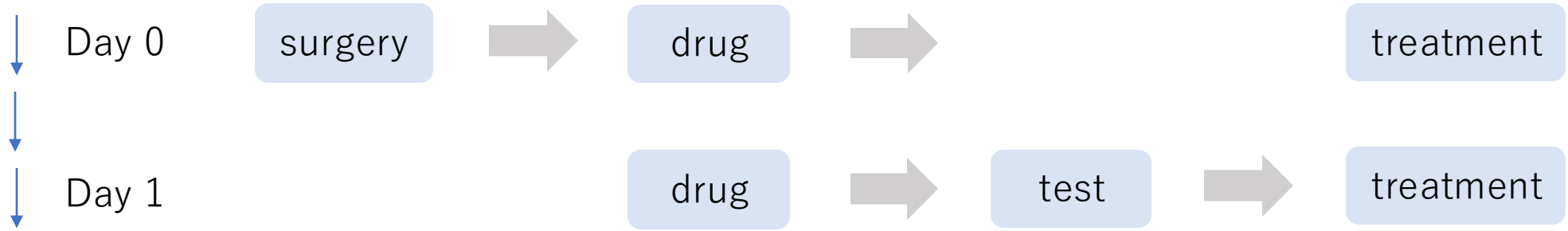
4

Step 4: Visualization

Extracted data is converted to JSON and visualized to clearly identify branching and differences in medical order sequences.

Step 1: Medical Order Sequence Construction

1. Extract medical orders, such as patient surgeries, medications, tests, and clinical procedures.
2. Assign the number of days elapsed from surgery to each medical order, using the surgery date as Day 0, and then categorize them by elapsed days.
3. For medical orders within the same elapsed days, reorder them by surgery → medication → test → treatment.

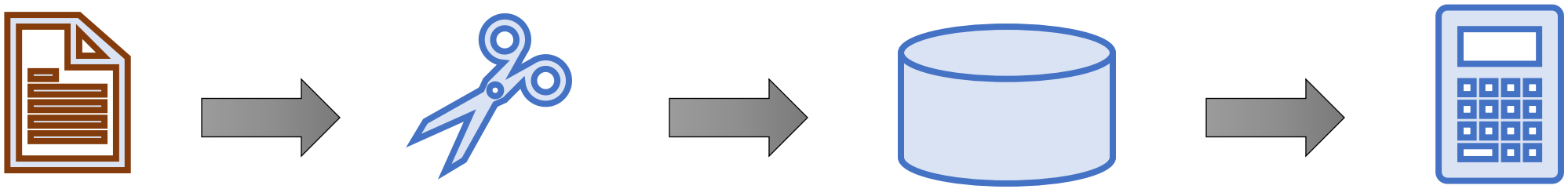


- A new table to store medical order sequence data will be created and added to the database.
- The new table schema will include SID data and be associated with medical order sequences for each patient.

Step 2:
Frequent Medical Order Pattern Extraction with Statistical Information

1. Frequent Medical Order Pattern Extraction

- Adopt T-PrefixSpan [1]
  - Extracts frequent patterns from sequences, taking into account time intervals.



Frequent Medical Order Pattern Extraction

Reduce Redundant Data (with Closed Patterns)

Retrieve Medication / Test Information

Extract Statistical Information

[1] **K. Uragaki, et al.** *Sequential Pattern Mining on Electronic Medical Records with Handling Time Intervals and the Efficacy of Medicines.* SoCC, 2016.

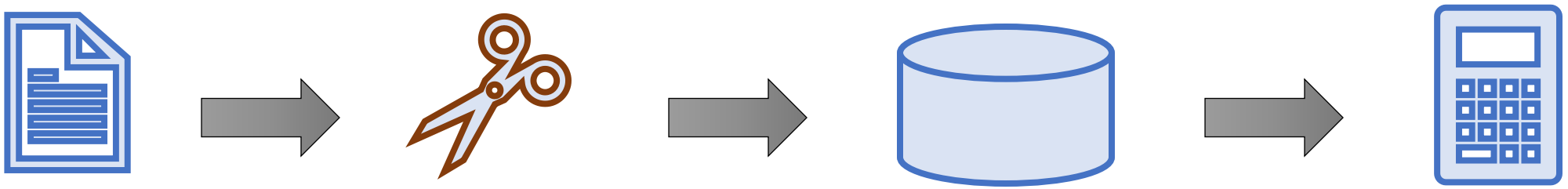
Step 2:
Frequent Medical Order Pattern Extraction with Statistical Information

2. Reducing Redundant Data with Closed Patterns

- By eliminating redundant data with closed patterns, we compact analysis results and improve efficiency.
- We'll extract only the patterns that don't have subsequences with higher occurrence.

Example of **Closed Patterns**

Pattern	Content	Occurrence
Pattern 1	Admission → Surgery → Test A	5
Pattern 2	**Admission → Surgery → Test A → Drug B**	**5**
Pattern 3	**Admission → Test C**	**7**
Pattern 4	**Admission → Test C → Drug D**	**4**



Frequent Medical Order Pattern Extraction

Reduce Redundant Data (with Closed Patterns)

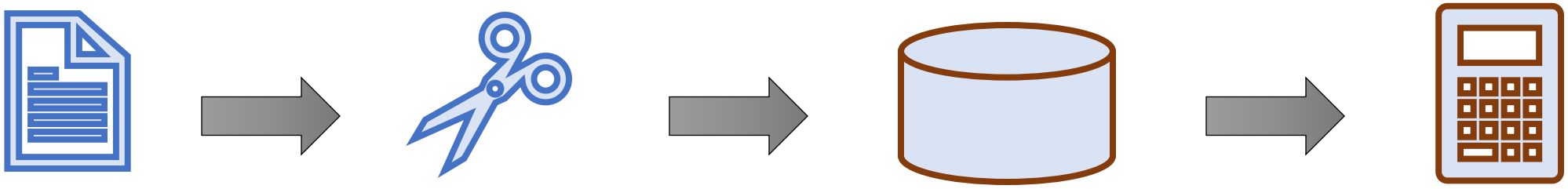
Retrieve Medication / Test Information

Extract Statistical Information

Step 2:
Frequent Medical Order Pattern Extraction with Statistical Information

3. Extracting Statistical Information from Frequent Medical Order Patterns

- Check if the frequent medical order pattern includes medication or test instructions.
- Search for corresponding patient information in the medication information and test information tables.
- Retrieve medication and test data for the relevant dates and calculate statistical information (e.g., mean, median, max, min).



Frequent Medical Order Pattern Extraction

Reduce Redundant Data (with Closed Patterns)

Retrieve Medication / Test Information

Extract Statistical Information

Step 3: Calculation of Abnormal Test Result Occurrence Rate

Calculate Abnormal Value Occurrence Rate

- Extract test results for each SID (Sequence ID, likely referring to individual patients or sequences) based on whether they are included in or excluded from the frequent medical order patterns.
- Aggregate abnormal values (high and low) for each elapsed day.
- Calculate the abnormal value occurrence rate for each test result.

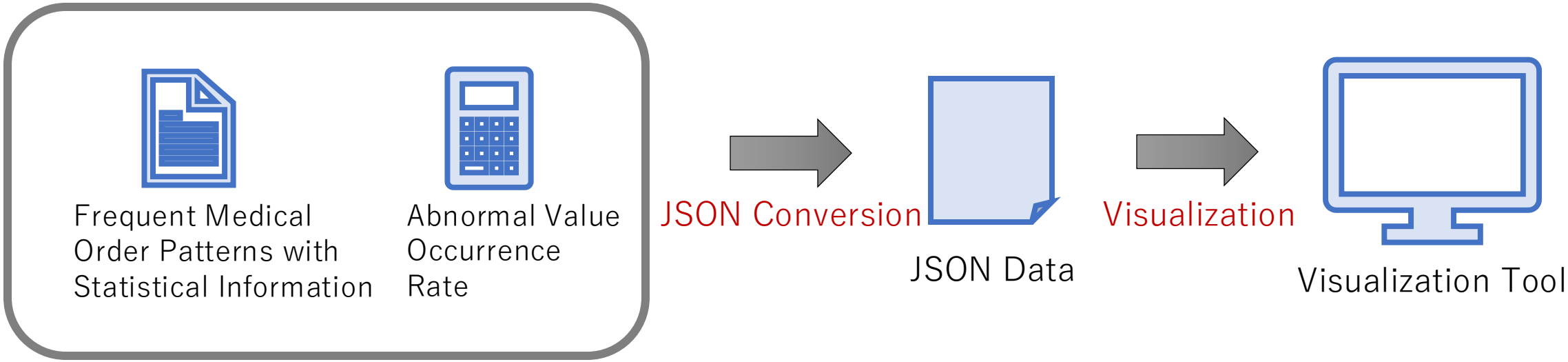
Definition of Normal and Abnormal Values

- Different reference ranges (normal ranges) are set for each test.
- Values exceeding or falling below the reference range are judged as "abnormal values”.

Step 4: Visualization

1. Visualization Method

- Convert the extracted frequent medical order patterns, statistical information, and abnormal value occurrence rates to JSON format.
- Load the data into a visualization tool to display the medical order patterns and abnormal value distributions for each disease.



Step 4: Visualization

2. Key Features of the Visualization Tool

- Frequent Medical Order Patterns & Statistical Information
  - Frequent medical order patterns are clearly visualized chronologically.
  - The placement of each pattern's node is determined by its support value.
  - Hovering over a node displays its associated statistical information in a tooltip.

- Abnormal Value Occurrence Rate
  - When a frequent medical order pattern is clicked, only patterns with the same support value are displayed.
  - Visualizes the proportion of "low, normal, and high values" for test results.
  - Data can be filtered by individual test item.
  - Hovering over a bar in the bar chart displays its statistical information in a tooltip.

Experimental Evaluation

Purpose

- To demonstrate the usefulness and feasibility of the proposed method.

Validation Items

1. Visualization of frequent instruction pattern variants.
2. Extraction of frequent instruction patterns with statistical information.
3. Calculation of normal and abnormal value occurrence rates in test results.

Dataset (1/2)

Real Data Overview

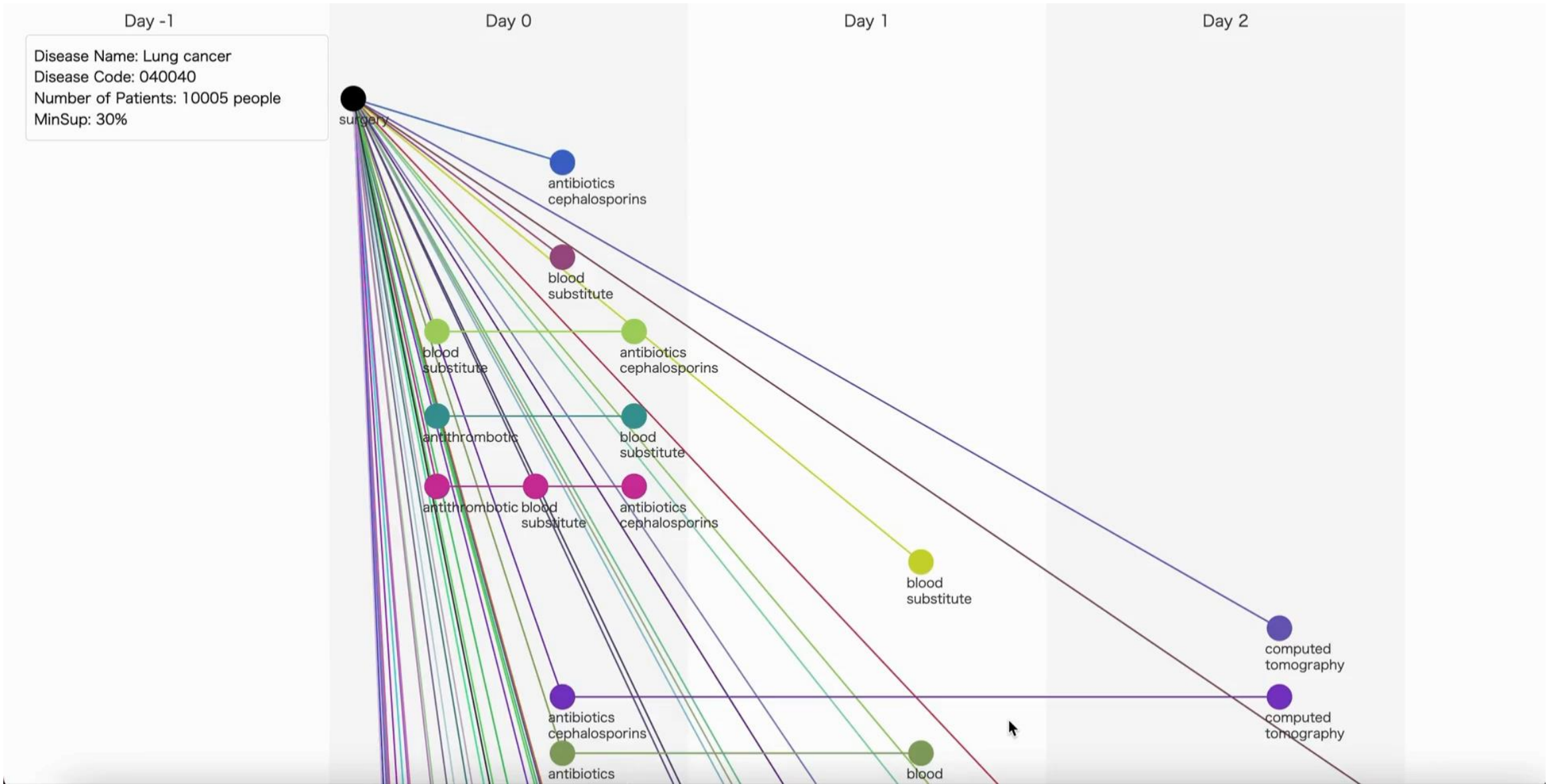
- Electronic health record (EHR) data from multiple medical institutions, anonymized and standardized for variations in notation/expression.
- 27 institutions, 9 diseases with 58,598 records
- 10 years (2015-2024)
- Reviewed and approved by the Utilization Purpose Review Committee of the Life Data Initiative.

Dataset (2/2)

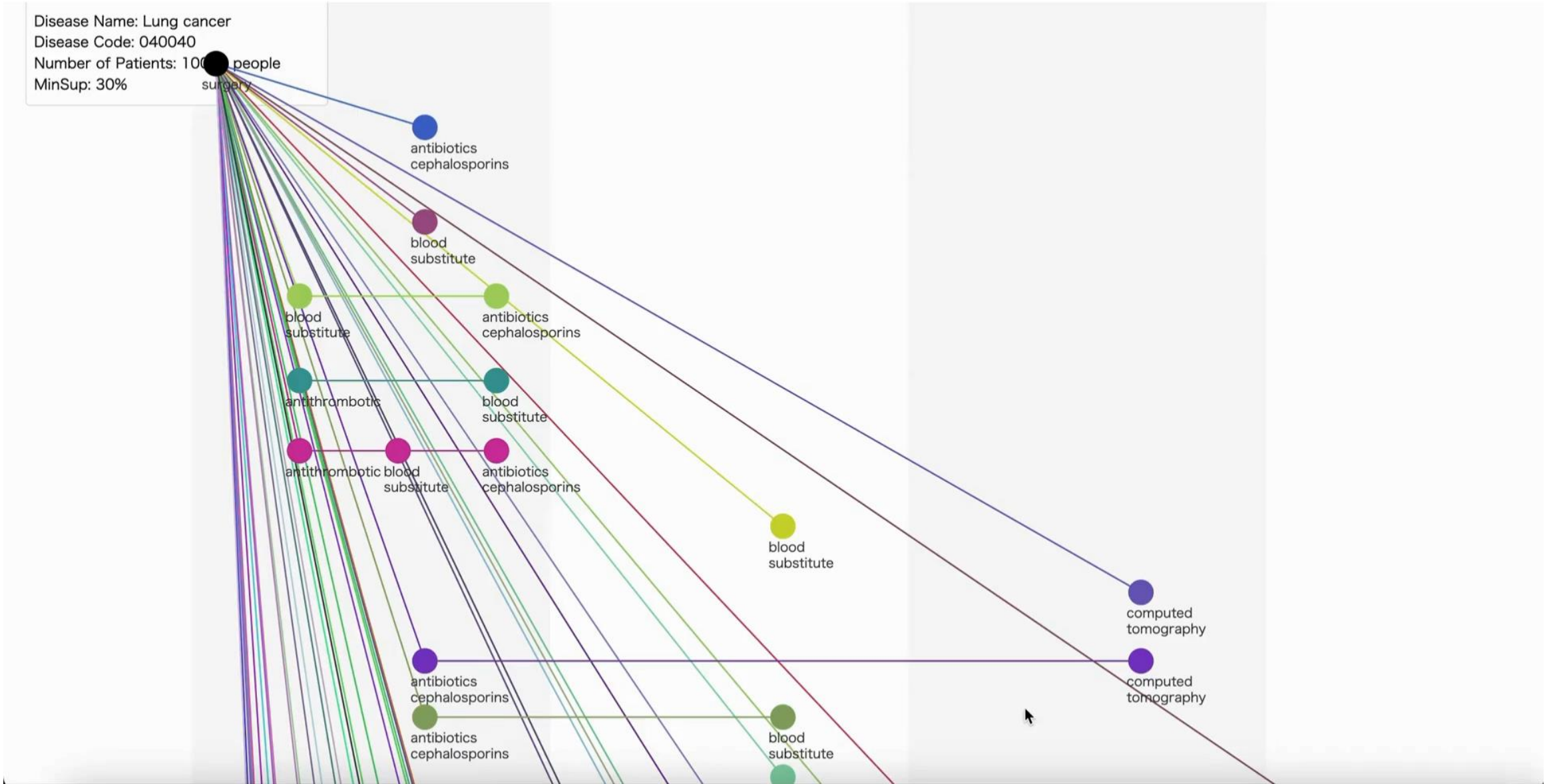
Patient Count by Disease

Disease	No. of Patients
Malignant lung tumors	9,962
Acute myocardial infarction	4,874
Angina pectoris and chronic ischemic heart disease	11,148
Malignant neoplasm of stomach	5,927
Malignant neoplasm of liver and intrahepatic bile ducts	2,030
Degeneration and hernia of intervertebral discs	1,194
Malignant neoplasm of breast	12,818
Bladder tumor	9,748
Malignant neoplasm of cervix and uterus	897
Total	58,598

Evaluation
Frequent Patterns Extraction



Evaluation
Abnormal Value Occurrence Rate



Conclusion and Future Work

Summary

- We extracted frequent medical order patterns with statistical information and abnormal test result occurrence rates.
- We built a system to visualize the extracted data.
- We re-validated our findings using real-world data and confirmed the reproducibility of the results.

Future Work

- Quantitative evaluation of frequent medical order pattern extraction results.
  - Comparison with typical clinical pathways.
- Estimating factors for branching and variants in frequent medical order patterns.
  - Analyzing differences in detail based on patient background (e.g., age, disease history).
  - Leveraging these insights for standardization and improvement.

Thank You

Statistical Analysis and Visualization of Medical Instruction Patterns in Multimedia-Rich Electronic Medical Records